Research Achievements

Dr. Shannon Quinn's research focus entails the discovery and development of rigorous spatiotemporal methods for characterizing biological motion in the context of induced perturbations: addition of a pathogen, contamination by a toxin, inclusion of a drug, or mutation of a gene. His main accomplishments are in the area of scalable algorithm development for processing large biomedical image and public health datasets. In short, Dr. Quinn studies "how organisms change and move" from a standpoint of computer vision and representation learning algorithms, and strategies for making these methods available to non-computational domain scientists.

Dr. Quinn is jointly appointed in the Departments of Computer Science (0.563) and Cellular Biology (0.187).

A. Publications and Presentations

Dr. Quinn has published 32 scientific articles in peer reviewed scientific journals and academic conferences, with 26 of these being published since the completion of his doctoral studies. Many of these publications have appeared in prestigious journals (*Science Translational Medicine*, *Journal of Immunology*, *IEEE Transactions in Big Data*) as well as some of the most selective international computer science, machine learning, and interdisciplinary computational science conferences (*IEEE BigData, IEEE ISBI, ACM SIGKDD, IEEE DSAA, SciPy*; all have acceptance rates under 20%). According to Google Scholar, his articles have been cited a total of 272 times since 2015.

Dr. Quinn has also been invited to present his research and share his expertise at departmental seminars (University of Georgia, University of Pittsburgh), local interest group meetings (PyData ATL), and regional conferences (Southern Data Science Conference, Biomedical Science and Engineering Conference, Jupyter Day @ Georgia Tech).

B. Funded Grants

Dr. Quinn has obtained extramural funding from the National Science Foundation (NSF) and the National Institutes of Health (NIH). From the NSF, Dr. Quinn has received three awards, two as PI: in 2015, a 3-year $768,834 Collaborative ABI-Innovations award ($90,000 to Dr. Quinn, as co-PI) with Dr. Chakra Chennubhotla (PI, University of Pittsburgh); in 2018, a 5-year $964,099 CAREER award, as PI; and in 2019, a 3-year $499,449 Innovations in Graduate Education (IGE) award, as PI. From the NIH, Dr. Quinn is co-I on a 2-year $429,807 award (Dr. Quinn has a total of 2.88 academic months and 2.0 summer months of effort) with Dr. Timothy Heckman (PI, University of Georgia). Dr. Quinn has also received two Faculty Research Grants from the University of Georgia, one in 2016 for $8,804, and one in 2017 for $9,829.

C. Advising

Dr. Quinn is currently a research advisor for four Ph.D. students (one in Infectious Diseases, three in Computer Science) and one Ph.D. candidate (Computer Science); one M.S. student (Computer Science); thirteen undergraduates (ten in Computer Science, four in Mathematics, two in Linguistics, two in Cognitive Science, and one in Economics; six of the thirteen are double majors); and one Young Dawgs high school student.

He has successfully graduated one Ph.D. student, nine M.S. students, six B.S. students, and one Young Dawgs (high school) student over the past five years, in addition to two REU students (summer 2015 and summer 2018).

He is on the advisory committee of ten graduate students at UGA.

D. Research Subareas

The percent of time assigned to research is 56% (0.42 research EFT out of 0.75 9-month academic year appointment).

Spatiotemporal Manifold Learning
Manifold learning is an area of machine learning concerned with deriving a low-dimensional representation of the original data that reveals more of the data's intrinsic structure, or to expose patterns in the data that are less apparent in the data's raw form. This makes learning manifolds of biological phenomena very attractive, given the extremely high dimensionality and low signal-to-noise ratio of biomedical image data. However, it also makes this approach very challenging: no "default" approach to successful manifold learning exists for a given biomedical problem. Furthermore, when incorporating temporal dynamics and evolving systems, no prescribed approach exists for learning a manifold over both spatial and temporal dimensions jointly. Dr. Quinn's lab has begun to explore the concept of "two-stream modeling," or jointly learning a decoupled spatial and temporal manifold, in the following biomedical applications.

Ciliary dyskinesia: Cilia are microscopic hairs that line the throat, nose, lungs, kidneys, brain, and numerous other organs in humans. Cilia are highly conserved organelles found throughout the animal kingdom. They beat in regular, rhythmic patterns, and can act as signaling centers. In multicellular organisms, they beat in unison to clear out particulates, irritants, and pathogens, and also play a critical role in early embryonic development in establishing signaling gradients for cell differentiation. As such, pathologies of the cilia—ciliopathies—have outsized negative impacts on the health, normal growth, and long-term maintenance of the host.

Unfortunately, because of the multi-organ nature of ciliopathies, they often present with nonspecific symptoms that are confused for other more common diseases. As most ciliopathies have genetic components and therefore require a genetic screen to positively identify, definitively diagnosing ciliopathies is challenging. Assessment of ciliary beat pattern—examination of the waveform generated by motile cilia—is generally accepted

as an effective tool for identifying and diagnosing ciliopathies. However, there is no quantitative “motion vocabulary” for ciliary dynamics: no common, validated methodology for evaluating ciliary beat pattern, and no established protocol for comparing waveform assessments across institutions or pathologies.

Dr. Quinn’s lab has made great strides in using computer vision to develop quantitative “elemental motion” descriptors of ciliary waveforms, and in forging the nascent theory of ciliary motion manifolds. Cilia have been shown to be well-described by “dynamic textures,” a fundamental element of computer vision theory characterized by rhythmic motions perturbed by inclusion of some random noise. These features showed remarkable robustness in automated binary classification, surpassing 90% accuracy when compared to human expert assessment of whether the observed ciliary motion in the videos constituted healthy or abnormal waveforms. This significant contribution was published in *Science Translational Medicine* and continues to be Dr. Quinn’s most-cited paper. Dr. Quinn’s lab has since leveraged these image invariants to train highly accurate segmentation procedures, whereby ciliated cells can be automatically identified in videos, akin to automatically identifying faces in photos.

For this work, Dr. Quinn was awarded a prestigious CAREER award from NSF for the next five years. One of the major aims of the project is to develop a rigorous manifold model, jointly capturing both ciliary appearance and dynamics, to fully enumerate the possible modes and pathologies of cilia waveforms. Dr. Quinn is approaching this challenging problem by leveraging Variational Autoencoders (VAEs), a powerful generative modeling approach to learning manifolds which implicitly account for sources of noise in the data while retaining high representation fidelity. Through a series of transformations, Dr. Quinn and his lab have devised a method in which the stationary components of the videos—appearance—are factored into one module of the VAE, while the changing components of the videos—dynamics—are factored into another. These two independent streams of information are then jointly projected onto the VAE manifold, representing cilia waveforms both in terms of their physical appearance (length, quantity, density) and their motion dynamics (beat frequency, coordination, stiffness).

Motility of Toxoplasma gondii *parasite*: *Toxoplasma (T.) gondii* is an intracellular parasite and the causative agent of disseminated toxoplasmosis. It is one of the world’s most successful parasites, capable of infecting virtually any warm-blooded animal, and is estimated to inhabit 30-50% of the world’s population. While the immune systems of most healthy adults can fight off *T. gondii* or keep it in a dormant state, those with suppressed or weakened immune systems are susceptible to serious infections and complications. *T. gondii* has been shown to induce behavior modifications in its hosts through a variety of subtle yet sophisticated epigenetic remodeling techniques. Its virulence, ultimately, is directly tied to its ability to invade host cells, reproduce, and “egress” the cells to look for more hosts to invade: in short, its lytic cycle, or ability to move. The *T. gondii* parasite is unique among the apicomplexa phylum for its motility mechanism, of which very little is known. Dr. Quinn’s lab has been working closely with Dr. Silvia Moreno of UGA to model the motility patterns of *T. gondii* in changing environments, to help address some underlying questions regarding *T. gondii* mechanobiology, in particular: what the

fundamental motion phenotypes are, how they are coordinated to guarantee a successful lytic cycle, and how they may indicate internal parasite state.

This has led to the development of numerous new approaches in both tracking and motility parameterization and evaluation. Dr. Quinn and his lab have confirmed, in numerous international and highly selective conferences (*IEEE ISBI, IEEE DSAA*, and *ACM SIGKDD*), the presence of three distinct *T. gondii* motility patterns and their connection to different intracellular concentrations of calcium, a universal cell signaling agent. Follow-up studies have confirmed the presence of these motility patterns even after accounting for different experimental conditions, providing strong evidence that these are fundamental motility patterns of the parasite that can be used to infer its internal state.

In object tracking, Dr. Quinn quickly discovered most off-the-shelf object trackers were either too brittle (sensitive to noise, especially in biological settings) or were fine-tuned to very specific applications, such as tracking cars on a highway, and not useful in a biomedical context. As such, he and his lab have developed a lightweight particle tracking method that is generic enough to be used in a variety of biomedical object tracking contexts, and which can also be parallelized to improve performance in large data contexts (this paper received Honorable Mention at *IEEE DSAA* 2019). This work has been incredibly successful, and there is an NIH R21 in progress to support further investigation into the connection between intracellular calcium concentration and *T. gondii* motility phenotype.

Subtyping progression of Parkinson's Disease: Parkinson's Disease (PD) is a neurodegenerative disorder characterized by neuronal degradation in the subcortical areas of the brain, resulting in progressively worsening motor, cognitive, and psychiatric symptoms. Following Alzheimer's Disease, PD is the most prevalent neurodegenerative disorder in the United States in adults over age 60. PD diagnosis, much like that of ciliopathy diagnosis, has traditionally relied on clinical assessments with a high degree of subjectivity. As a result, early-stage PD is often missed altogether, and quantitative benchmarks for subtyping PD stage and progression, or even differentiating PD from other neurodegenerative disorders, are sparse to nonexistent.

Given the relatively new availability of open source neuroimage datasets, Dr. Quinn is uniquely positioned, in collaboration with his faculty mentor Dr. Tianming Liu, to leverage his lab's big data and biomedical imaging expertise and Dr. Liu's neuroscience expertise to explore the problem of identifying subtypes of PD progression and methods for delineating these subtypes diagnostically. NIH has identified PD subtyping as one of its priority research areas in neurodegeneration, and Dr. Quinn is currently working with Dr. Liu and neuroscientists at University of Pittsburgh on an NIH K-25 award, which would afford Dr. Quinn the flexibility to undergo basic neuroscience training while also leveraging his existing strengths in biomedical imaging, machine learning, and big data to more thoroughly address this research question. Already, Dr. Quinn's lab has published two conference papers to *SciPy* on different methodologies of modeling longitudinal, multi-modal brain image data to capture PD subtypes and progression. In addition, Drs.

Liu and Quinn were co-senior authors on a landmark 2016 paper to the prestigious *ACM SIGKDD* conference on how these sophisticated brain imaging techniques could be bootstrapped into big data processing frameworks.

Graph theoretic analysis of organellar networks: Motivated by the observation in Fine-Coulson *et al* 2015 that infection of A549 human type II epithelial cells with *Mycobacterium tuberculosis* (Mtb) induces changes in mitochondrial morphology, spatial distribution, and mass that are dependent on a specific active gene in the pathogen, Dr. Quinn has focused on developing imaging and modeling methods that can capture the spatiotemporal dynamics of diffuse subcellular structures, e.g. mitochondria or actin cytoskeletal proteins. These structures are difficult to quantify with off-the-shelf bioimaging algorithms that tend to focus more on closed structures like the cell nucleus or cell wall.

Dr. Quinn was co-PI on NSF award #1458766 to develop these imaging algorithms from a graph theoretic perspective: by considering these diffuse subcellular structures as *loosely-coupled dynamic social networks*, rather than rigid geometric structures, changes in the shape, quantity, and spatial distribution of these structures could be tracked over time and used in model inference for genetic screens such as in the motivating example of mitochondrial degradation in the presence of Mtb. Preliminary findings were published in 2018 at the *SciPy* conference, and the initial software package release was only recently accepted to the *Journal of Open Source Software* in 2020 for inspection and application by researchers across the world.

Biosurveillance

Dr. Quinn has been studying machine learning methods for observing large quantities of “non-traditional” data streams and making intelligent inferences about the potential for emerging public health threats, with the ultimate goal of developing a means for mobilizing resources ahead of an outbreak to mitigate or otherwise minimize its effects. Of course, this work has taken on a new relevance with COVID-19, but was originally motivated by the 2009 H1N1 pandemic: most biosurveillance approaches were limited, either in using “official” data outlets such as the CDC (which are reliable, but delayed), or in using “non-traditional” data sources such as Twitter (which are real-time, but notoriously unreliable).

Instead, in collaboration with colleagues at the Georgia Department of Public Health, Oak Ridge National Laboratory, and the UGA School of Public Health, Dr. Quinn has been working on leveraging the strengths of both data streams—the reliability of public health experts, with the spatial and temporal resolution of social media—to develop a biosurveillance platform that could detect public health threats and help mobilize responses before the threats could rise to pandemic levels. Originally termed “ORBiT” (Oak Ridge Biosurveillance Toolkit) and with a demonstrated ability to show both spatial and temporal spread effects of the 2009 H1N1 pandemic from social media data, Dr. Quinn has since made additional contributions in the development of *temporal word embeddings*: the ability to track unique key words and phrases as they appear and disappear (e.g., hashtags) without manual intervention. These contributions resulted in NIH award #1R21DA047893-01A1 to conduct a retrospective study on the 2015-2016

Scott County, Indiana HIV outbreak, and to determine if, had a biosurveillance platform like ours been in operation at the time, the outbreak could have been recognized sooner.

Open and Reproducible Science

Dr. Quinn has been an active participant in and contributor to the Open Science movement since becoming an Apache Mahout project committer in 2011. Open Science is an interdisciplinary, rigorous paradigm for making all research materials, methods, data, and artifacts universally accessible for inspection and replication. Adoption and inculcation of Open Science tools and practices often have high "activation costs," as many of these practices contravene existing institutional and academic incentives. Nevertheless, as open source and open data have revolutionized the fields of artificial intelligence and machine learning, Dr. Quinn has worked to reduce barriers to access and adoption of Open Science tools and techniques.

Since 2016, Dr. Quinn has organized and chaired the Open Science in Big Data (OSBD) Workshop, held annually in conjunction with the *IEEE BigData Conference*, a very prestigious international venue for interdisciplinary data and computational science. Each year's proceedings have featured different perspectives on the challenges and rewards of Open Science adoption, particularly within the context of "big data." The workshop attracted guest speakers from HHMI Janelia Farms, Georgia Tech, University of Washington, New York University, Google, Oak Ridge National Laboratory, Cray, and Lucidworks. Acceptance rate decreased to 50% in its final year, its selectivity record for the workshop.

Most of Dr. Quinn's publications are accompanied by availability of code for the purpose of study replication; recently, Dr. Quinn had a code module accepted for publication by the Journal of Open Source Software (JOSS), which conducts extensive peer review of the software for both methodology and coding standards compliance. In 2016, Dr. Quinn was invited to submit a manuscript on open source "big data" frameworks and their strengths and weaknesses, given his expertise in both big data and open science. He has been invited to give numerous guest lectures across UGA campus on Open Science and its applications in different fields. He was recently recognized for his work in furthering Open Science by NSF with award #1955049, a 3-year pilot program in graduate education, specifically for introducing a scalable and maintainable interdisciplinary Open Science curriculum.

Dr. Quinn is on the Editorial Board for the Journal of Open Source Education (JOSE).